Asian Seabass and arowana genome Analyses

Stephen J OBrien

NSU-OC Jamboree

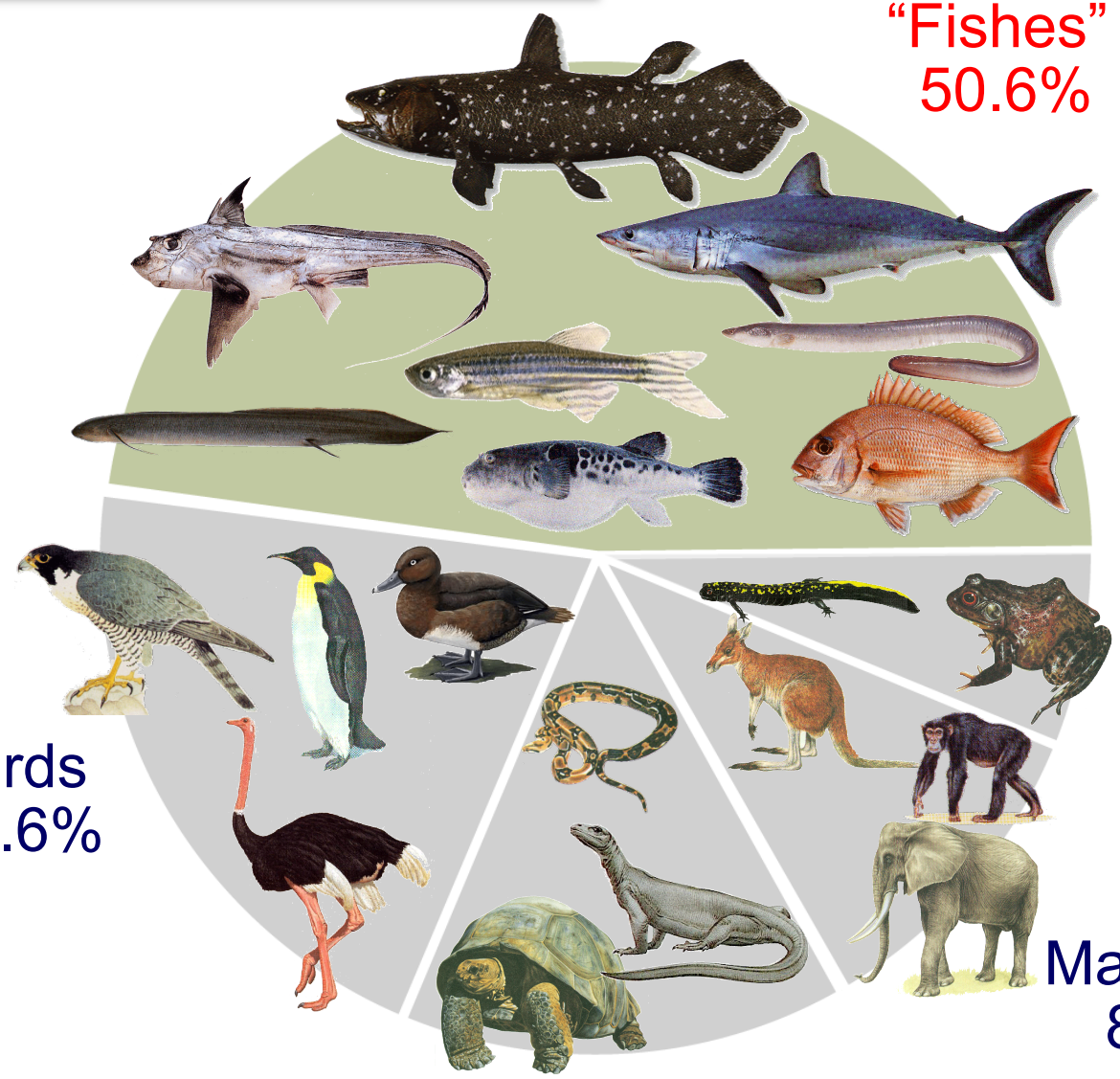
May 19 2016

10,000 vertebrate genomes

GENOME 10K©

Molecular menagerie. These animals and thousands of other vertebrates may one day have their genomes sequenced.

Vertebrates ~ 62,000 species



Examples of Selected Species

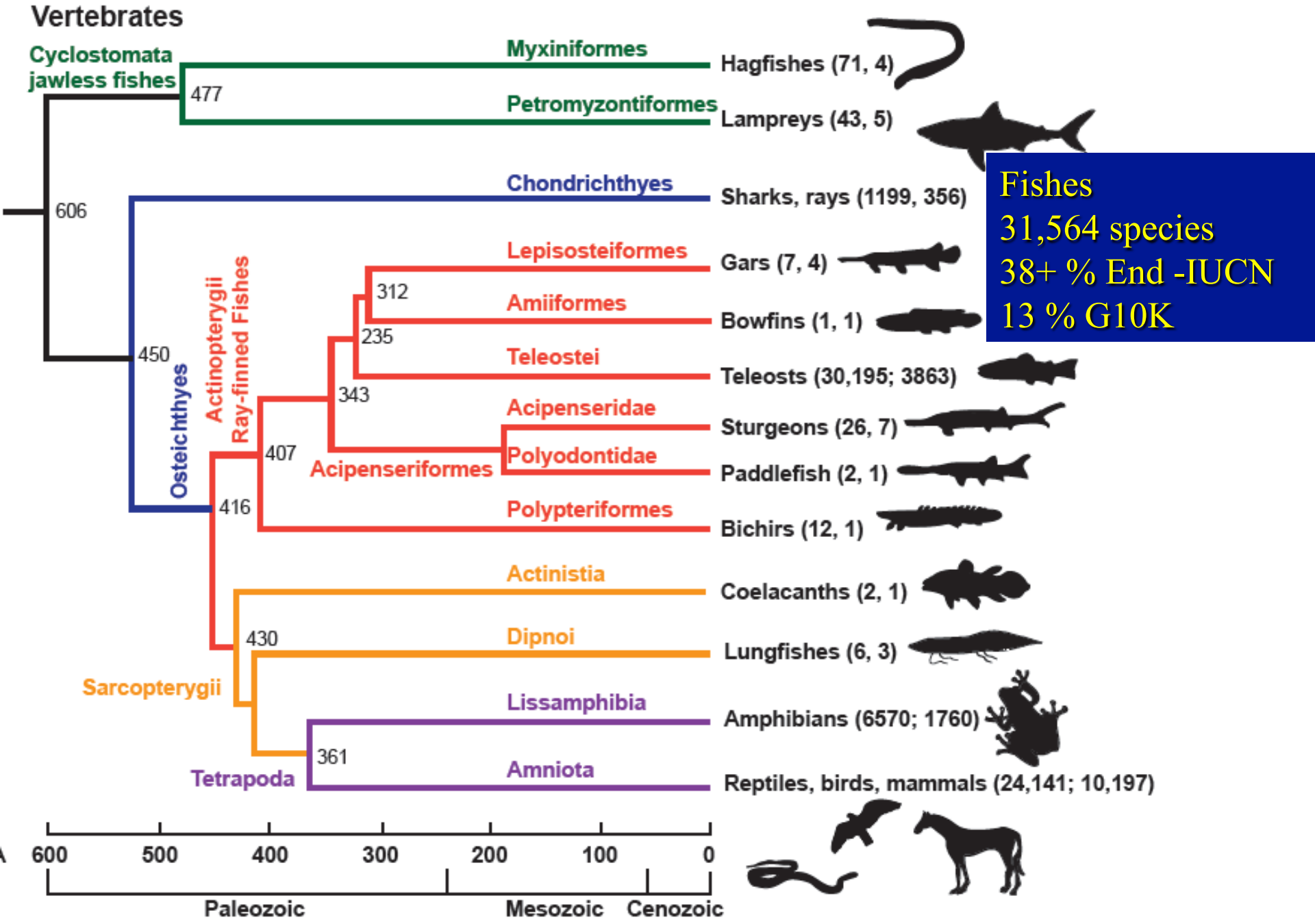
Horned marsupial frog

Monito del Monte

GENOME 10K©

The most diverse group of vertebrates 31,564 species

Modified from Meyer, 2005


Vertebrates
Cyclostomata
jawless fishes
Myxiniformes
Hagfishes (71, 4)
477
Petromyzontiformes
Lampreys (43, 5)
Chondrichthyes
Sharks, rays (1199, 356)
606
Lepisosteiformes
Gars (7, 4)
312
Amiiformes
Bowfins (1, 1)
235
Teleostei
Teleosts (30,195; 3863)
450
Osteichthyes
Actinopterygii
Ray-finned Fishes
343
Acipenseridae
Sturgeons (26, 7)
407
Acipenseriformes
Polyodontidae
Paddlefish (2, 1)
416
Polypteriformes
Bichirs (12, 1)
Actinistia
Coelacanths (2, 1)
430
Dipnoi
Lungfishes (6, 3)
Sarcopterygii
Lissamphibia
Amphibians (6570; 1760)
361
Tetrapoda
Amniota
Reptiles, birds, mammals (24,141; 10,197)
600
500
400
300
200
100
0
Paleozoic
Mesozoic
Cenozoic
Fishes
31,564 species
38+ % End -IUCN
13 % G10K

PLOS | GENETICS

RESEARCH ARTICLE

Chromosomal-Level Assembly of the Asian Seabass Genome Using Long Sequence Reads and Multi-layered Scaffolding

Shubha Vij[1]*, Heiner Kuhl[2], Inna S. Kuznetsova[1,3], Aleksey Komissarov[4], Andrey A. Yurchenko[4], Peter Van Heusden[5], Siddharth Singh[6], Natascha M. Thevasagayam[1], Sai Rama Sridatta Prakki[1], Kathiresan Purushothaman[1], Jolly M. Saju[1], Junhui Jiang[1], Stanley Kimbung Mbandi[5], Mario Jonas[5], Amy Hin Yan Tong[7], Sarah Mwangi[5], Doreen Lau[1], Si Yan Ngoh[1], Woei Chang Liew[1], Xueyan Shen[1], Lawrence S. Hon[6], James P. Drake[6], Matthew Boitano[6], Richard Hall[6], Chen-Shan Chin[6], Ramkumar Lachumanan[6], Jonas Korlach[6], Vladimir Trifonov[8], Marsel Kabilov[9], Alexey Tupikin[9], Darrell Green[10], Simon Moxon[11], Tyler Garvin[12], Fritz J. Sedlazeck[12,13], Gregory W. Vurture[12], Gopikrishna Gopalapillai[14], Vinaya Kumar Katneni[14], Tansyn H. Noble[15], Vinod Scaria[16], Sridhar Sivasubbu[16], Dean R. Jerry[15], Stephen J. O'Brien[4,17], Michael C. Schatz[12,13], Tamás Dalmay[18], Stephen W. Turner[6], Si Lok[19], Alan Christoffels[5]*, László Orbán[1,20,21]*

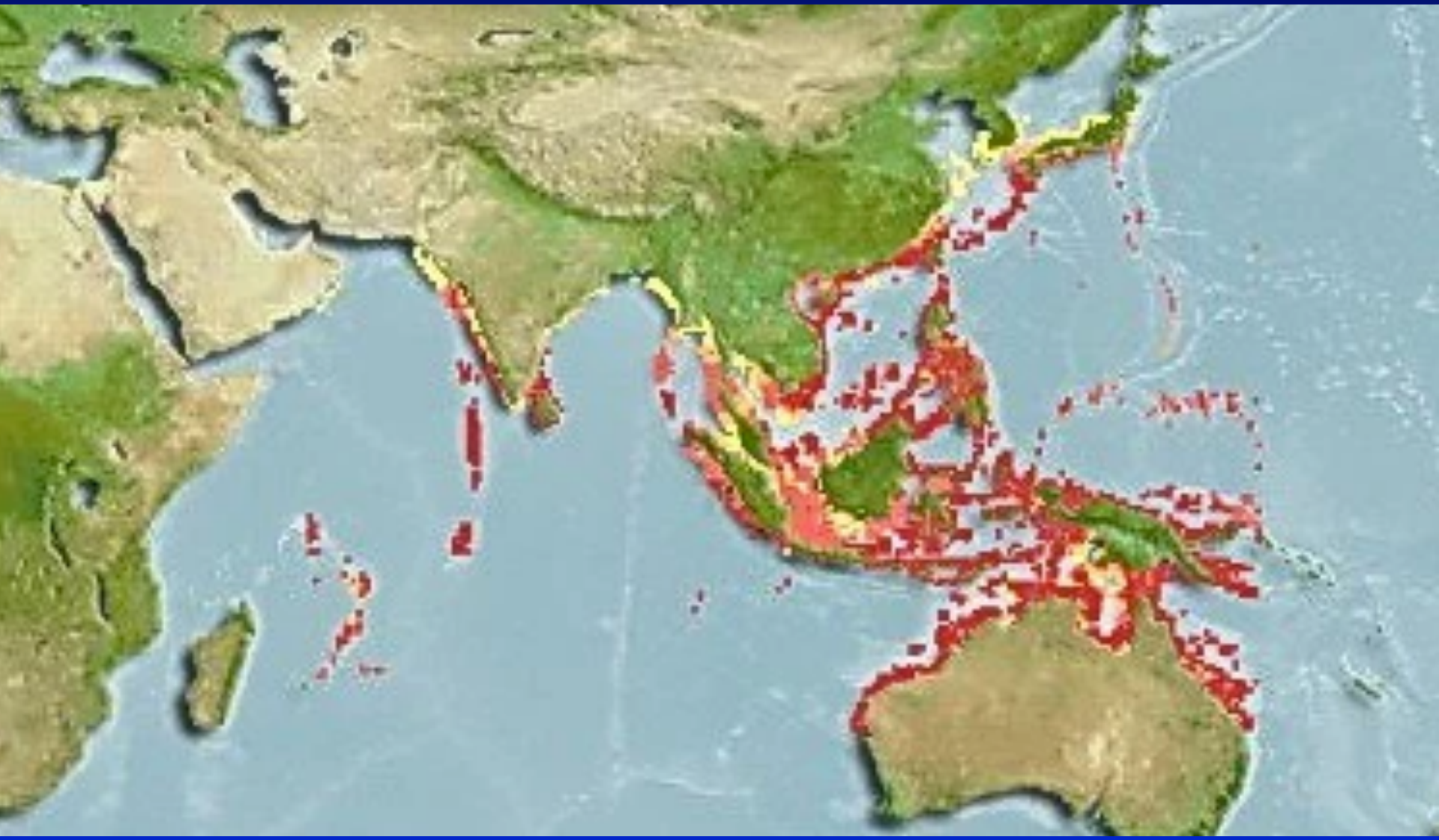


Table 1. Assembly and scaffolding statistics for the Asian seabass genome.

Primary Genome Assembly (v1)		
	Number of contigs	3,917
	Contig N50/count	1,066,117/139
	Max. contig size	18,910,200
	Total size	668,453,369
Scaffolded Genome Assembly (v2)		
	Number of scaffolds	3,807
	Scaffold N50/count	1,191,366/119
	Max. scaffold size	18,910,200
	Total size	668,464,831
Chromosome-Level Genome Assembly (v3)		
	Number of chromosomes	24
	Scaffold N50/count	25,848,596/11
	Max. scaffold size	30,776,907
	Total size	586,924,032

Promising mapping and long read technologies

- Pacific bioscience
- Oxford Nanopore
- Illumina Moleculo
- Life Technologies
- Complete Genomics
- OpGen
- BioNanoGenomics
- Trios
- RH maps
- Sperm haplotyping

N50 contigs – 1Mbp
N50 scaffolds 25Mbp

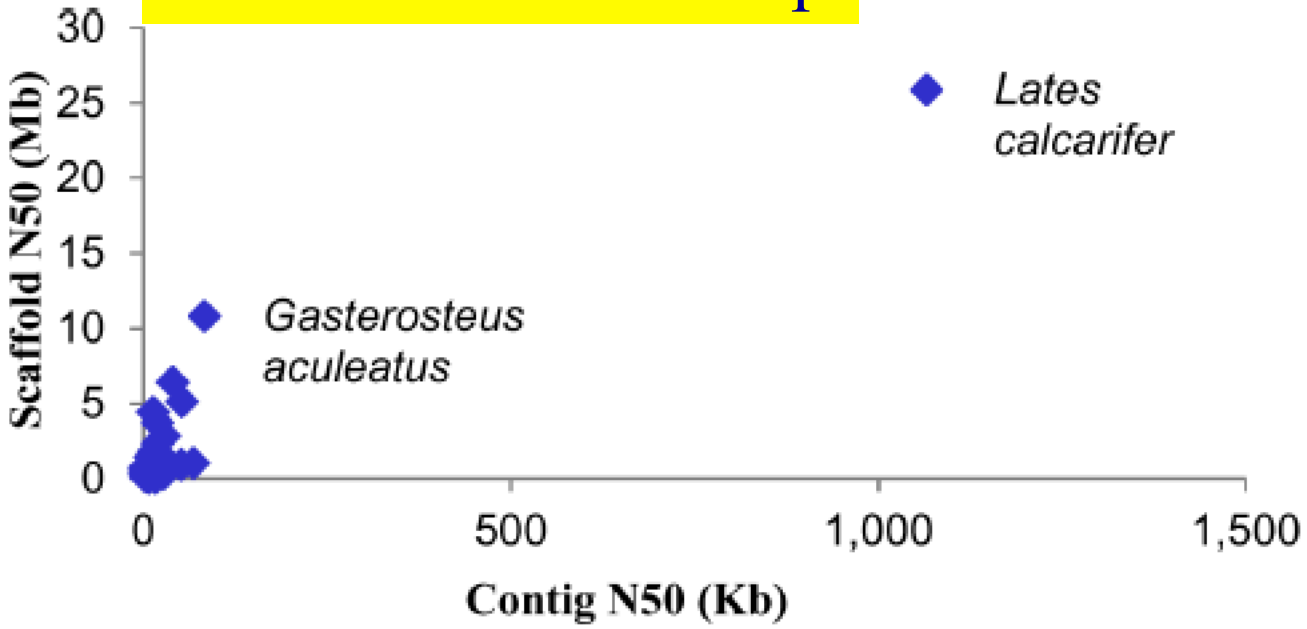
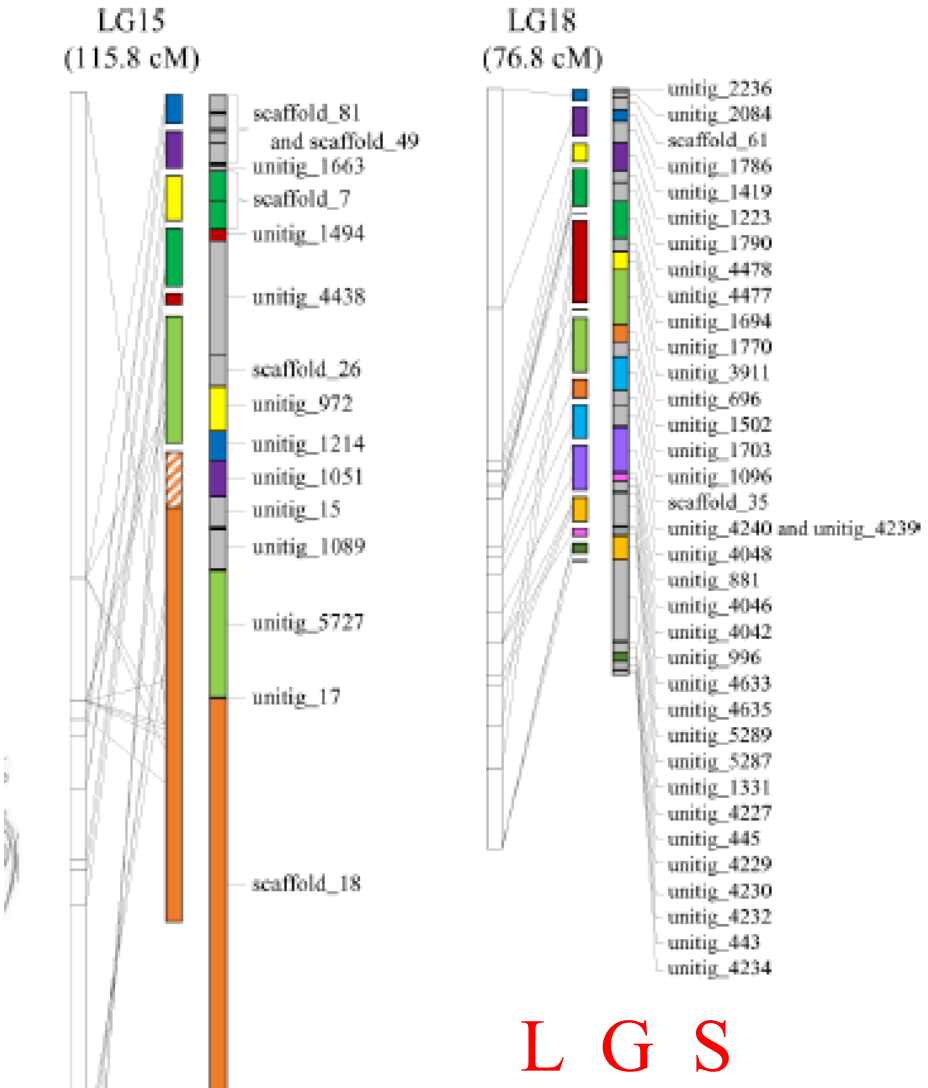


Fig 1. *Lates calcarifer* has the best metrics from among the assembled fish genomes till date. The *L. calcarifer* genome contig N50 and scaffold N50 values were compared to the following fish genomes: *Anguilla japonica*, *Astatotilapia burtoni*, *Astyanax mexicanus*, *Boleophthalmus pectinirostris*, *Ctenopharyngodon idellus*, *Cynoglossus semilaevis*, *Cyprinus carpio*, *Danio rerio*, *Dicentrarchus labrax*, *Electrophorus electricus*, *Esox lucius*, *Gadus morhua*, *Gasterosteus aculeatus*, *Larimichthys crocea*, *Latimeria chalumnae*, *Metriaclima zebra*, *Neolamprologus brichardi*, *Notothenia coriiceps*, *Oncorhynchus mykiss*, *Oreochromis niloticus*, *Oryzias latipes*, *Pundamilia nyererei*, *Periophthalmodon schlosseri*, *Periophthalmus magnuspinnatus*, *Salmo salar*, *Scartelaos histophorus*, *Takifugu flavidus*, *Takifugu rubripes*, *Tetraodon nigroviridis*, *Thunnus*


LG15
(115.8 cM)
scaffold_81
and scaffold_49
unitig_1663
scaffold_7
unitig_1494
unitig_4438
scaffold_26
unitig_972
unitig_1214
unitig_1051
unitig_15
unitig_1089
unitig_5727
unitig_17
scaffold_18
LG18
(76.8 cM)
unitig_2236
unitig_2084
scaffold_61
unitig_1786
unitig_1419
unitig_1223
unitig_1790
unitig_4478
unitig_4477
unitig_1694
unitig_1770
unitig_3911
unitig_696
unitig_1502
unitig_1703
unitig_1096
scaffold_35
unitig_4240 and unitig_4239
unitig_4048
unitig_881
unitig_4046
unitig_4042
unitig_996
unitig_4633
unitig_4635
unitig_5289
unitig_5287
unitig_1331
unitig_4227
unitig_445
unitig_4229
unitig_4230
unitig_4232
unitig_443
unitig_4234

L G S

Genome Annotation

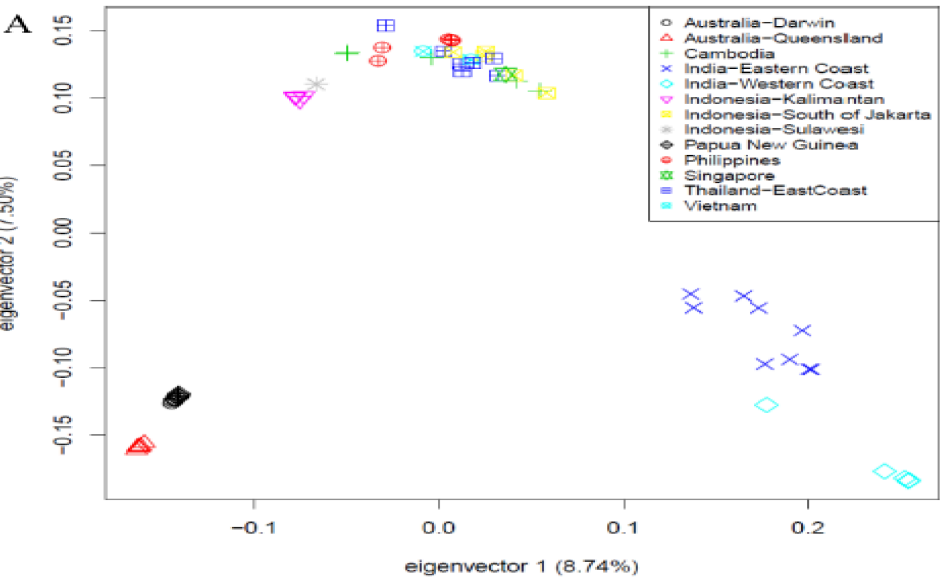
Annotated genome features

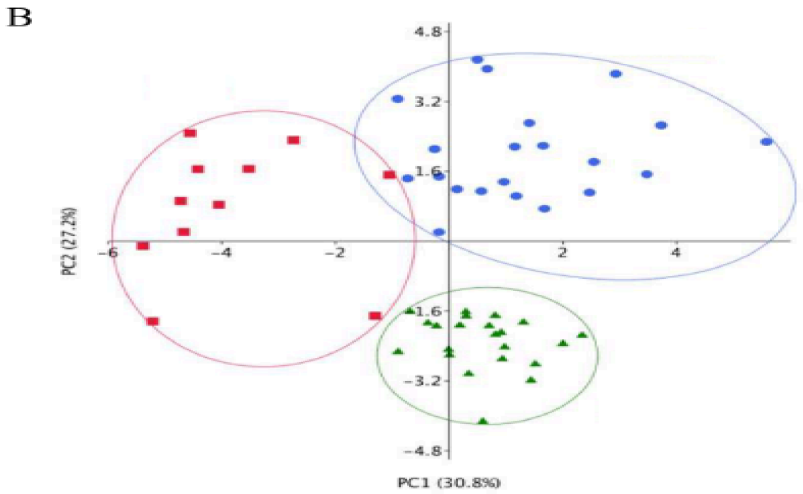
- Feature
- I. Assembly Statistics
- II. Genome Browser
- III. Gene annotation
- IV. DNA variants
- a. SNPs
- b. Indels
- c. Copy number variation
- V. Repeats content
- a. Interspersed repeats
- b. Tandem repeats
- · Complex Tandem Repeats
- · Short Tandem Repeats
- · STR primer sequences
- VI. Evolutionary Constrained Elements (ECE) among mammals
- VII. Endogenous retrovirus-like elements
- VIII. Methylation Sites in the cat genome
- IX. Micro-RNA
- X. Nuclear mitochondrial segments (Numts)
- XI. Segmental duplications –cop number variation
- XII. Gene expansion/contraction
- XII. Signatures of Selection
- a. Ds/Dn ratio
- b. Long Fst
- c. Long homozygosity
- d. Long haplotypes

Table 2. Annotation statistics of the Asian seabass genome.

Annotation	
Protein-coding genes	22,184
Mean transcript length (bp)	13,448
Mean coding DNA sequence length (bp)	1,737
Mean exons/gene	10
Mean exon length (bp)	170
Mean intron length (bp)	11,714
rRNAs	1,828
miRNAs	3,024
tRNAs	2,077
snoRNAs	212
snRNAs	1,153
miscRNAs	209

doi:10.1371/journal.pgen.1005954.t002

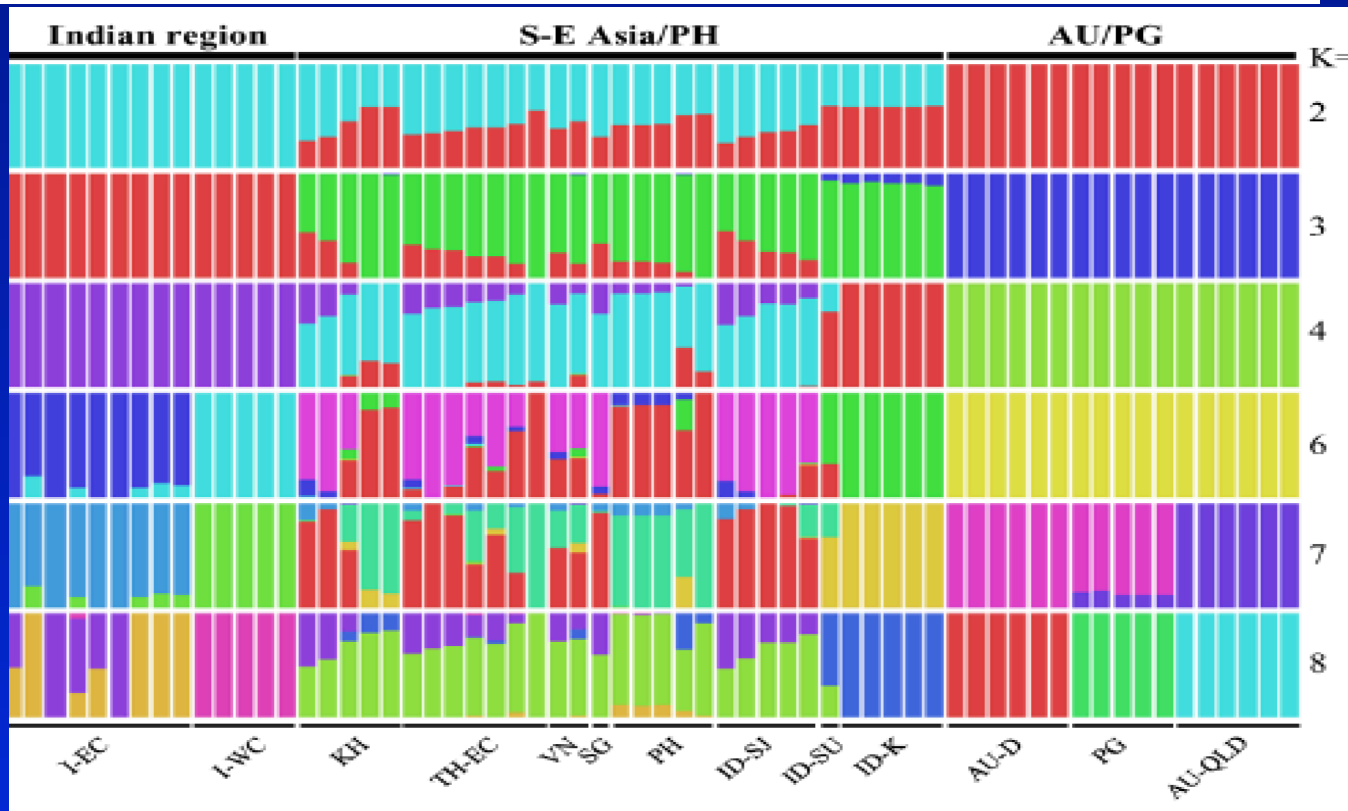




PCA-WGS and Morph above

61 WGS

STRUCTURE below



SCIENTIFIC REPORTS

OPEN

Received: 13 November 2015
Accepted: 30 March 2016
Published: 19 April 2016

The Asian arowana (*Scleropages formosus*) genome provides new insights into the evolution of an early lineage of teleosts

Chao Bian[1,2,*], Yinchang Hu[3,*], Vydianathan Ravi[4,*], Inna S. Kuznetsova[5,6,*], Xueyan Shen[5,*], Xidong Mu[3,*], Ying Sun[2], Xinxin You[1,2], Jia Li[1,2], Xiaofeng Li[7], Ying Qiu[1,2], Boon-Hui Tay[4], Natascha May Thevasagayam[5], Aleksey S. Komissarov[8], Vladimir Trifonov[9,10], Marsel Kabilov[11], Alexey Tupikin[11], Jianren Luo[3], Yi Liu[3], Hongmei Song[3], Chao Liu[3], Xuejie Wang[3], Dangen Gu[3], Yexin Yang[3], Wujiao Li[2], Gianluca Polgar[12], Guangyi Fan[2], Peng Zeng[2], He Zhang[2], Zijun Xiong[2], Zhujing Tang[2], Chao Peng[1,2], Zhiqiang Ruan[1,2], Hui Yu[1,2], Jieming Chen[1,2], Mingjun Fan[1,2], Yu Huang[1,2], Min Wang[1,2], Xiaomeng Zhao[1,2], Guojun Hu[1,2], Huanming Yang[2,13,14], Jian Wang[2,13], Jun Wang[2,13,15], Xun Xu[2], Linsheng Song[16], Ganqchun Xu[17], Pao Xu[17], Junmin Xu[2,18], Stephen J. O'Brien[8,19], László Orbán[5,20,21], Byr

GBE

Whole Genome Sequencing of the Asian Arowana (*Scleropages formosus*) Provides Insights into the Evolution of Ray-Finned Fishes

Christopher M. Austin[1,2,†], Mun Hua Tan[1,2,†], Larry J. Croft[1,2,3], Michael P. Hammer[4], and Han Ming Gan[1,2,*]

Native

Introduced

Myanmar

Laos

Thailand

Vietnam

Cambodia

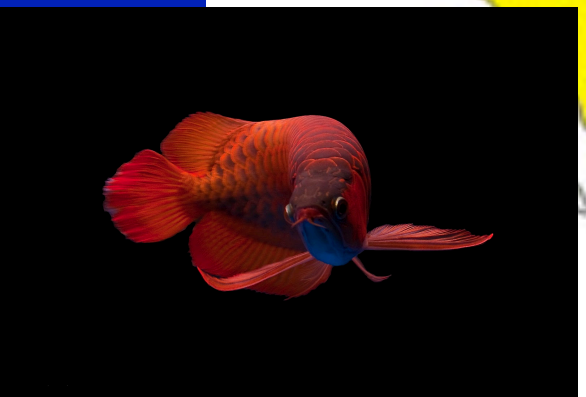
Philippines

Brunei

Malaysia

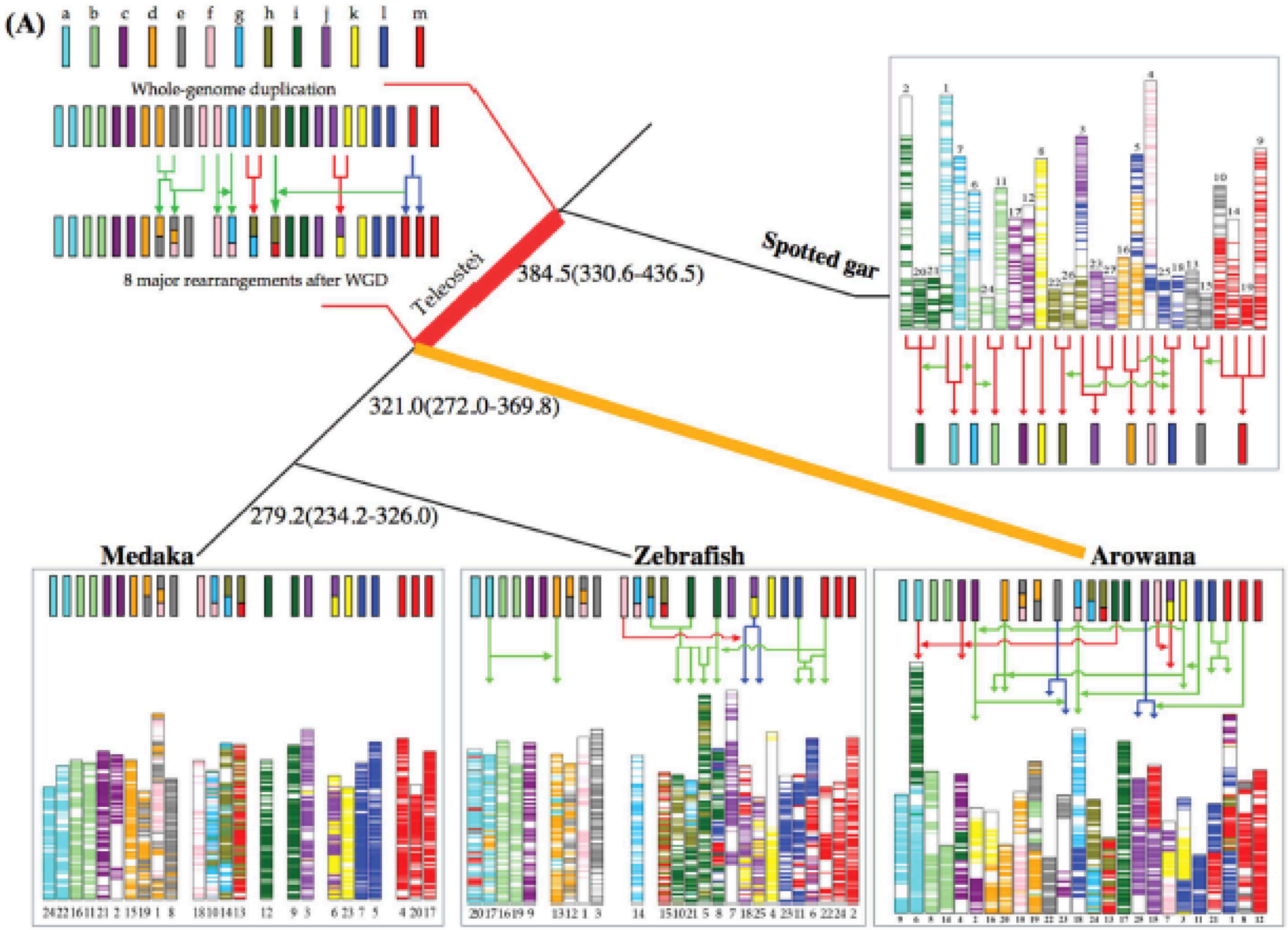
Singapore

Indonesia

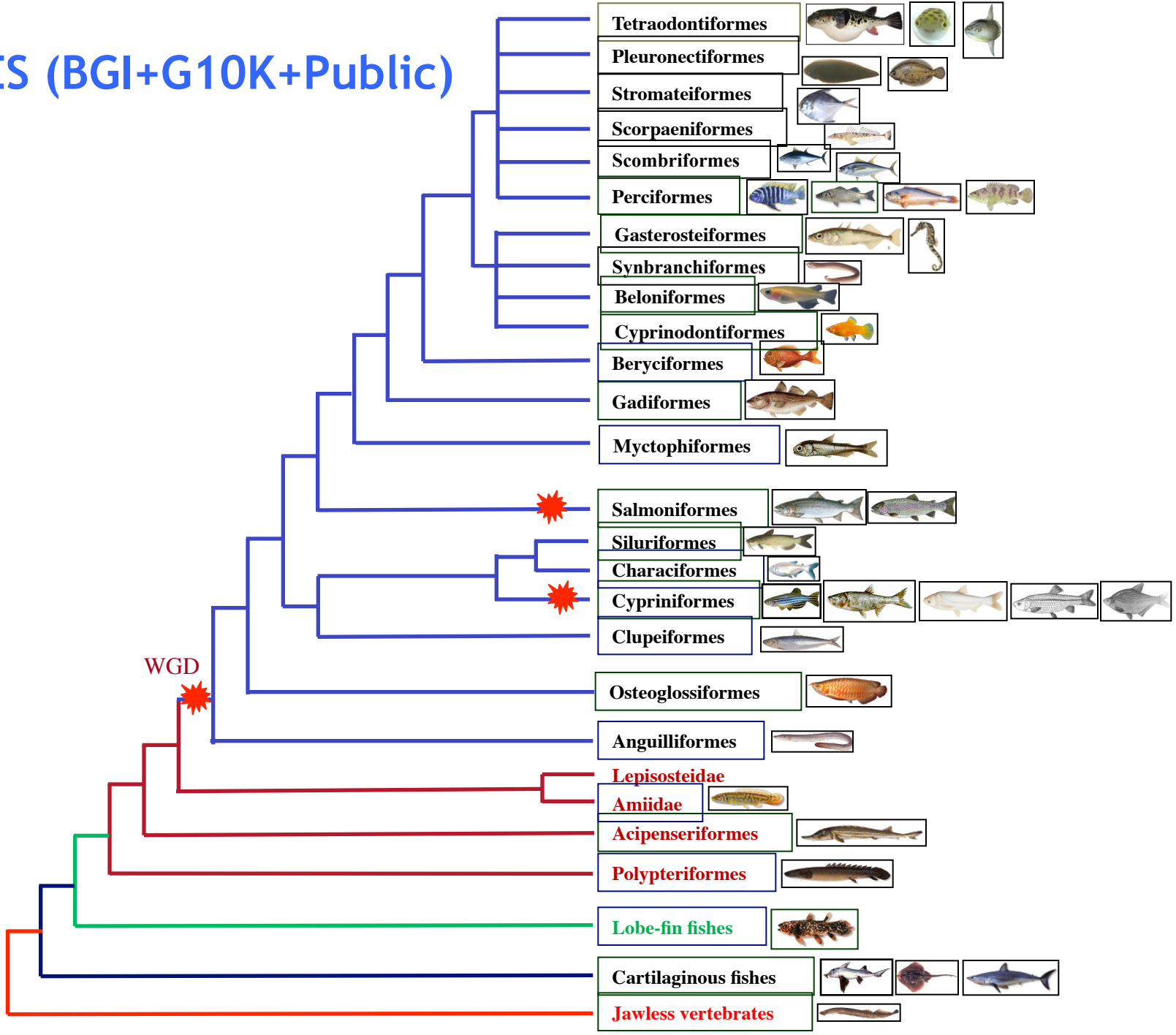


Colour variety	Golden	Red	Green
Sequence coverage (-fold)	138	110	100
Estimated genome size (Gb)	0.82	0.95	0.90
Assembled genome size (Gb)	0.78	0.75	0.76
Scaffold N50 (Mb)	5.97	1.63	1.85
Contig N50 (kb)	30.73	60.19	62.80
Number of genes	22,016	21,256	21,524
Repeat content	27%	28%	28%

Table 1. Overview of the genome assembly and annotation


(A)
a b c d e f g h i j k l m
Whole-genome duplication
8 major rearrangements after WGD
Teleostei
384.5(330.6-436.5)
Spotted gar
321.0(272.0-369.8)
279.2(234.2-326.0)
Medaka
Zebrafish
Arowana

FISHES (BGI+G10K+Public)



Theodosius Dobzhansky Center for Genome Bioinformatics St. Petersburg

St. Petersburg State University

Founded 1724